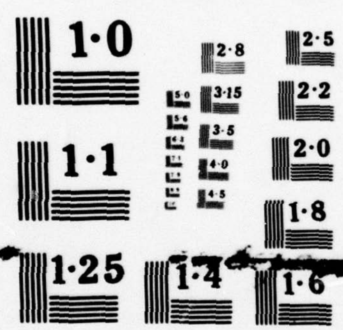
AD-A069 820 UNION COLL AND UNIV SCHENECTADY NY INST OF ADMINISTR--ETC F/G 12/1
A USEFUL APPROXIMATE FORMULA FOR THE DETERMINATION OF THE REGIO--ETC(U)
FEB 79 L A AROIAN, R W MILLER N00014-77-C-0438

UNCLASSIFIED AES-7901 NL

1 OF 1
AD
A069820

END
DATE
FILMED
7-79
DDC



NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

A USEFUL APPROXIMATE FORMULA FOR THE DETERMINATION OF THE REGIONS OF THE SEQUENTIAL TEST FOR THE CORRELATION COEFFICIENT

AES-7901

FOUNDED IN 1795
UNION COLLEGE
SCHENECTADY N.Y.

ADMINISTRATIVE AND ENGINEERING SYSTEMS MONOGRAPH

(14) AES-79Ø1

(6) A USEFUL APPROXIMATE FORMULA FOR THE DETERMINATION OF THE REGIONS OF THE SEQUENTIAL TEST FOR THE CORRELATION COEFFICIENT.

(9) Technical rept.
Jun 78-Jan 79, by

(10) Leo A./Aroian
Union ~~College and Uni~~versity

Robert W./Miller
Statistical Scientist
Allied Chemical Company, Buffalo, N.Y.

(11) 15 Feb[redacted] 1979 (12) 23p.

(15) NØØØ14-77-C-Ø438

UNION COLLEGE AND UNIVERSITY
Institute of Administration and Management
Schenectady, New York 12308

408 725 LB

A USEFUL APPROXIMATE FORMULA FOR THE DETERMINATION
OF THE REGIONS OF THE SEQUENTIAL TEST
FOR THE CORRELATION COEFFICIENT

by

Leo A. Aroian
Union College and University

Robert W. Miller
Statistical Scientist
Allied Chemical Company, Buffalo, N.Y.

1. Introduction.

A simple approximate formula is shown to be remarkably accurate for the determination of the regions of the sequential test for the correlation coefficient, ρ, when the variates follow a bivariate normal distribution. The approximate results are compared with the exact values and with an approximation of Ghosh (1970). The exact results depend on the solution of an equation involving the ratio of two hypergeometric functions. These results, combined with the Monte Carlo evaluations of the OC, operating characteristic function, the ASN, the average sample number, and the exact values of some selected regions by Campbell, Taneja, and Aroian (1977), (1978) provide all the materials needed for this sequential test. The real advantage of a sequential test over a fixed size test is the early termination of the test when ρ exceeds ρ_1 or ρ is less than ρ_0 by a substantial amount. Additionally the savings in sample size vary from 25% to 80% as compared to a fixed size test.

2. Determination of the Regions.

The sequential test for the coefficient of correlation is due to B.K. Ghosh (1970), who gives a complete discussion of all its essential properties. The test is described as follows. Let

$$\bar{x}_{1n} = \sum_{i=1}^{n} x_{1i}/n, \ \bar{x}_{2n} = \sum_{i=1}^{n} x_{2i}/n,$$

$$s_{1n}^2 = \sum_{i=1}^{n} (x_{1i}-\bar{x}_{1n})^2/n, \ s_{2n}^2 = \sum_{i=1}^{n} (x_{2i}-\bar{x}_{2n})^2/n,$$

$$r_n = \sum_{i=1}^{n} (x_{1i}-\bar{x}_{1n})(x_{2i}-\bar{x}_{2n})/ns_{1n}s_{2n},$$

where $ns_{1n}s_{2n} = \{\sum(x_{1i}-\bar{x}_{1n})^2 \ \sum(x_{2i}-\bar{x}_{2n})^2\}^{1/2}$

The hypothesis under test is H_0: $\rho = \rho_0$ versus H_1: $\rho = \rho_1$, $\rho_1 > \rho_0$. The Wald sequential test limits are given by $r_n(u)$, the upper limit for r_n, and $r_n(\ell)$, the lower limit for r_n. As soon as $r_n \le r_n(\ell)$ accept $\rho = \rho_0$, or as soon as $r_n \ge r_n(u)$ accept $\rho = \rho_1$. The values of $r_n(u)$ and $r_n(\ell)$ are determined as follows. First $z_n(r_n)$ must be found:

$$z_2(r_2) = \ell n(\pi-2\sin^{-1}\rho_1) - \ell n(\pi-2\sin^{-1}\rho_0), \text{ if } r_2=-1,$$

$$= \ell n(\pi+2\sin^{-1}\rho_1) - \ell n(\pi+2\sin^{-1}\rho_0), \text{ if } r_2=1,$$

$$\begin{aligned} z_n(r_n) = {} & .5(n-1)\{\ell n(1-\rho_1^2) - \ell n(1-\rho_0^2)\} \\ & -(n-1.5)\{\ell n(1-\rho_1 r_n) - \ell n(1-\rho_0 r_n)\} \\ & +\ell n\ F\{.5,.5,n-.5;\ .5(1+\rho_1 r_n)\} \\ & -\ell n\ F\{.5,.5,n-.5;\ .5(1+\rho_0 r_n)\}, \text{ if } n > 2. \end{aligned}$$

Note $F(r,s,t;z) = \sum_{j=0}^{\infty} \frac{\Gamma(r+j)\Gamma(s+j)\Gamma(t)}{\Gamma(r)\Gamma(s)\Gamma(t+j)} (z^j/j!)$.

Define $b = \ln\{\beta/(1-\alpha)\}$, $a = \ln\{(1-\beta)/\alpha\}$. If $n = 2$, $r_2 = -1$, and $z_2(-1) \le b$, accept $\rho = \rho_0$; if $n = 2$, $r_2 = 1$, and $z_2(1) \ge a$, accept $\rho = \rho_1$. If $n \ge 3$, r_n is computed from $z_n(r_n) = b$ and $z_n(r_n) = a$, and $\rho = \rho_0$ or $\rho = \rho_1$ is accepted depending on whether $r_n \le r_n(\ell)$, or $r_n \ge r_n(u)$ where $r_n(\ell)$ and $r_n(u)$ are solutions of $z_n(r_n(\ell)) = b$, and $z_n(r_n(u)) = a$. This test, the preceding results, and the monotonicity properties of $z_n(r_n)$ are due to B.K. Ghosh (1970).

3. Approximate Formula for r_n.

If we omit the hypergeometric functions in $z_n(r_n) = a$ or b we obtain the approximate formula for $r_n(u)$ and $r_n(\ell)$ for $n \ge 3$:

(3.1) $r_n(u)$ or $r_n(\ell) = (W-1)/(W\rho_1-\rho_0)$

(3.2) $W = \{(1-\rho_0^2)/(1-\rho_1^2)\}^{1/2+(4n-6)^{-1}} e^{\omega(n-1.5)^{-1}}$

where $\omega = b$ for determining $r_n(u)$, and $\omega = a$ for determining $r_n(\ell)$. The formula is correct to order $O(n^{-1})$ and is exact if the hypergeometric terms could be neglected. The hypergeometric terms depend only on $r = 0.5$, $s = 0.5$, $t = n-0.5$, $z = 0.5(1+\rho_1 r_n)$ and $0.5(1+\rho_0 r_n)$. *A remainder term* C_1 is found if we use $\ln(1+A_1 z)$ for each hypergeometric function

(3.3) $C_1 = 0.5(\rho_1-\rho_0)(4n-2)^{-1} r_n$,

bounded by ± 0.1 for $n > 2$ in the worst possible situation $\rho_1 = 1$, $\rho_0 = -1$ and $r_n = \pm 1$; is least when $\rho_1 - \rho_0$ and r_n are small. Furthermore a and b are the smallest in absolute

value when α and β assume their largest permissible values. Hence the approximate formula will be quite accurate for α's and β's commonly used: $\alpha \leq .25$, $\beta \leq .25$. A more accurate remainder term C_2 of order $O(n^{-3})$ is found by use of three terms of the hypergeometric function and its logarithm:

$$(3.4) \quad C_2 = 0.5(\rho_1-\rho_0)r_n\{(4n-2)^{-1} + (1.125(4n^2-1)^{-1} - 0.5(4n-2)^{-2})(1+0.5(\rho_1+\rho_0)r_n)\}.$$

Here r_n indicates either $r_n(\ell)$ or $r_n(u)$. Both C_1 and C_2 will be used to improve the accuracy of (3.1) and (3.2).

A second approximation to $r_n(\ell)$ and $r_n(u)$ is first to calculate

$$(3.5) \quad W_1 = W \exp\{\omega_1(n-1.5)^{-1}\}$$

where $\omega_1 = b-C_1$ for $r_n(u)$ and $\omega_1 = -a+C_1$ for $r_n(\ell)$. A third approximation is to use $\omega_1 = b-C_2$ and $-a+C_2$ in W_1. In both C_1 and C_2 the r_n to be used is that value found first by (3.1) and (3.2). Otherwise we may consider r_n unknown in C_1 and C_2, and then solve for it by using $z_n(r_n(\ell)) = a$ and b, where the hypergeometric terms are properly replaced by C_1 or C_2.

4. Examination of Various Cases.

We examine a case where (3.1) and (3.2) are almost as poor as possible, Table 1, and a more usual set of values, Table 2. Let $\rho_0 = -.5$, $\rho_1 = .5$, $\alpha = .40$, $\beta = .50$, $a = .22314$, $b = -.18232$. In this case the results are given in Table 1. Clearly formulas (3.1) and (3.2) either alone, or followed by (3.5) using

C_1 or C_2 are certainly sufficient. In Table 2 we list the exact boundaries values, those found by the approximation (3.1), (3.2), and Ghosh's approximation formula (1970, p. 324) for a wide variety of $\{\rho_0, \rho_1, \alpha, \beta\}$.

TABLE 1

Boundaries for the Sequential Test

$\alpha = .40$, $\beta = .50$, $\rho_1 = .50$, $\rho_0 = -.50$

	$r_n(u)$				$r_n(\ell)$			
n	1	2	3	4	5	6	7	8
3	.1485	.1436	.1421	.1418	-.1214	-.1174	-.1163	-.1160
4	.0892	.0879	.0876	.0875	-.0729	-.0719	-.0716	-.0716
5	.0637	.0632	.0631	.0631	-.0521	-.0517	-.0516	-.0516
6	.0496	.0493	.0493	.0493	-.0405	-.0403	-.0403	-.0403
7	.0406	.0404	.0404	.0404	-.0332	-.0330	-.0330	-.0330

1 & 5 evaluated by (3.1) and (3.2),

2 & 6 evaluated by the second approximation (3.5),

3 & 7 evaluated by the third approximation (3.5),

4 & 8 are exact results.

Corrections to AES-7901

A USEFUL APPROXIMATE FORMULA FOR THE DETERMINATION OF THE REGIONS OF THE SEQUENTIAL TEST FOR THE CORRELATION COEFFICIENT

In Table 2

					$r_n(\ell)$			$r_n(u)$		
ρ_0	ρ_1	α	β	n	1	2	3	4	5	6
0	.10	.05	.05	30			-.9273			
				31	-.9928	-.9933	-.8957	.9957	.9961	.9944
				40	-.7396	-.7400	-.6829	.7831	.7833	.7819

TABLE 2

COMPARISONS OF APPROXIMATE AND
EXACT VALUES OF r_n

ρ_0	ρ_1	α	β	n	$r_n(\ell)$			$r_n(u)$		
					1	2	3	4	5	6
0	.10	.05	.05	30			-.9273			
				31		-.9933	-.8957		.9961	.9944
				40		-.7400	-.6829		.7833	.7819
0	.10	.05	.15	20	-.9925	-.9923				.9672
				30	-.6143	-.6147	-.8904	.9922	.9925	.6615
				40	-.4375	-.4378	-.6553	.7564	.7566	.5086
0	.10	.10	.15	20	-.9604	-.9603				.9403
				30	-.5941	-.5946	-.6605	.7703	.7707	.6436
				40	-.4229	-.4232	-.4828	.5885	.5887	.4952
0	.10	.20	.01	18			-.8354	.9698	.9707	
				20			-.7468	.8747	.8754	
				30			-.4812	.5936	.5940	
				46	-.9787	-.9789	-.2964	.4017	.4019	.9974
0	.10	.20	.05	18			-.8126	.9473	.9481	
				20			-.7263	.8543	.8551	
				30	-.9650	-.9655	-.4675	.5800	.5803	.9689
				40	-.6919	-.6922	-.3381	.4452	.4454	.7392
0	.10	.20	.10	17			-.8317	.9709	.9717	
				20			-.6994	.8276	.8283	
				30	-.7016	-.7020	-.4496	.5621	.5624	.7390
				40	-.5012	-.5014	-.3246	.4317	.4320	.5668
0	.10	.20	.15	16			-.8512	.9956	.9966	
				20	-.8899	-.8907	-.6710	.7993	.7999	.8817
				30	-.5505	-.5509	-.4306	.5432	.5435	.6045
				40	-.3911	-.3914	-.3104	.4175	.4178	.4659
0	.10	.20	.20	16	-.9449	-.9463	-.8127	.9579	.9589	.9114
				20	-.7219	-.7227	-.6402	.7690	.7697	.7392
				30	-.4446	-.4449	-.4101	.5230	.5234	.5095
				40	-.3138	-.3140	-.2950	.4024	.4026	.3946
0	.25	.05	.10	11	-.8984	-.9003	-.8963			.9140
				20	-.3701	-.3704	-.4358	.6898	.6901	.5599
				30	-.1889	-.1890	-.2482	.5023	.5025	.4156
				40	-.1044	-.1045	-.1544	.4086	.4086	.3444
.25	.50	.05	.10	9	-.9865	-.9913	-.5762			.5832
				10	-.7480	-.7508	-.4804			.5630
				15	-.2206	-.2211	-.1929	.8769	.8773	.4486
				20	-.0285	-.0286	-.0492	.7668	.7670	.4725
.25	.50	.10	.10	9	-.9338	-.9383	-.3508			
				10	-.7071	-.7097	-.2775	.9547	.9555	
				20	-.0167	-.0168	.0522	.6889	.6891	.6954
				30	.1382	.1382	.1621	.5914	.5916	.5909

TABLE 2 (Continued)

ρ_0	ρ_1	α	β	n	$r_n(\ell)$ 1	$r_n(\ell)$ 2	$r_n(\ell)$ 3	$r_n(u)$ 4	$r_n(u)$ 5	$r_n(u)$ 6
.25	.50	.10	.20	7	-.7775	-.7840	-.5109			.9876
				10	-.2569	-.2579	-.2430	.9313	.9331	.8059
				20	.1268	.1269	.0695	.6750	.6752	.5939
				30	.2226	.2227	.1736	.5815	.5816	.5233
.25	.75	.10	.20	4	-.3279	-.3367	-.0307			.8814
				10	.3678	.3681	.3160	.7698	.7703	.6808
				15	.4396	.4398	.3930	.6976	.6978	.6362
.25	.75	.20	.20	4	-.2097	-.2152	.1457	.9609	.9656	.8514
				10	.3858	.3861	.3865	.7107	.7113	.6688
				15	.4499	.4501	.4400	.6555	.6558	.6282

1 and 4 are exact results,

2 and 5 are new approximations,

3 and 6 Ghosh's approximations.

A blank indicates no decision.

In all cases the boundaries found by the new approximation are wider than those given by the exact results. This means that the ASN (average sample number) will be slightly larger, with an accompanying slight decrease in α and β.

REFERENCES

Campbell, D., Taneja, V., Aroian, L.A. (1977). Sequential tests for the coefficient of correlation. Exact Wald regions, operating characteristic, and average sample number. 1977 Social Statistics Section Proceedings of the American Stat. Asso., 987-990.

Ghosh, B.K. (1970). Sequential Tests of Statistical Hypotheses. Addison-Wesley.

Taneja, V., Campbell, D., Aroian, L.A. (1978). Tables of the regions, operating characteristic function, and average sample number for Wald sequential tests, Ohio State Univ. Technical Report 158, Dept. of Stat., 1958 Neil Avenue, Columbus, Ohio, 43210.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AES-7901	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Useful Approximate Formula for the Determination of the Regions of the Sequential Test for the Correlation Coefficient		5. TYPE OF REPORT & PERIOD COVERED Technical Report-- Jun. 1978-Jan.-1979
		6. PERFORMING ORG. REPORT NUMBER AES-7901
7. AUTHOR(s) Leo A. Aroian Robert W. Miller		8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0438
9. PERFORMING ORGANIZATION NAME AND ADDRESS Institute of Administration & Management, Union College Schenectady, NY 12308		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of Navy Arlington, VA		12. REPORT DATE Feb. 15, 1979
		13. NUMBER OF PAGES 9
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Presented to the Computing Section of the American Stat. Assoc. at its annual meeting, August 1979 at Washington, D.C.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Sequential analysis, coefficient of correlation. boundary regions, approximate formula, Wald regions.

rho

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A simple approximate formula is shown to be remarkably accurate for the determination of the regions of the sequential test for the correlation coefficient, ρ, when the variates follow a bivariate normal distribution. The approximate results are compared with the exact values and with an approximation of Ghosh (1970). The exact results depend on the solution of an equation involving the ratio of two hypergeometric functions. (over)

next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20 (continued)

These results, combined with the Monte Carlo evaluations of the OC, operating characteristic function, the ASN, the average sample number, and the exact values of some selected regions by Campbell, Taneja, and Aroian (1977), (1978) provide all the materials needed for this sequential test. The real advantage of a sequential test over a fixed size test is the early termination of the test when ρ exceeds ρ_1 or ρ is less than ρ_0 by a substantial amount. Additionally the savings in sample size vary from 25% to 80% as compared to a fixed size test.

rho rho

rho sub 1

rho sub 0